

Application du système GenFam à la réponse au stress des plantes : intégration de l'identification d'éléments cis spécifiques

Jonathan LORENZO^{1,2}, Delphine LARIVIERE², Gaëtan DROC², Jean-Francois DUFAYARD², Dominique THIS³ and Stéphanie BOCS²

¹ Master 2 STIC pour la santé, Faculté des Sciences, Place Eugène Bataillon, 34095 Montpellier cedex 5, France

² CIRAD, UMR AGAP, Montpellier F-34398, France

³ Montpellier SupAgro, UMR AGAP, Montpellier F-34060, France

Corresponding author: jonath.lorenzo@gmail.com

Résumé

GenFam est un système intégratif d'analyse de familles de gènes. Ce système permet (i) de créer des familles de gènes de génomes complets, (ii) d'exécuter une analyse phylogénétique de cette famille à travers le gestionnaire de workflows Galaxy afin de définir les relations d'homologie, (iii) d'étudier des événements évolutifs à partir de blocs de synténie précalculées avec le workflow SynMap de la plateforme de génomique comparative (CoGe) et (iv) d'intégrer ces résultats dans l'interface de visualisation synthétique. La première application de GenFam est d'identifier des gènes candidats pour la tolérance aux stress environnementaux. Il nécessite de mettre en évidence la présence de séquences régulatrices cis spécifiques de la réponse aux stress (de type ABRE, DRE). Dans ce contexte, nous avons besoin d'intégrer de nouveaux outils afin de découvrir et chercher des sites de fixation de facteurs de transcription (Transcription Factor Binding Sites, TFBS) dans les séquences promotrices des gènes membre de la famille étudiée. Ce workflow Galaxy va, d'une part, sélectionner les régions flanquantes en 5' ou en 3' des gènes d'intérêts selon le choix de l'utilisateur. D'autre part, les régions flanquantes sont analysées afin de découvrir et rechercher les motifs de séquences régulatrices cis spécifiques de la réponse aux stress avec des méthodes complémentaires comme MEME, STIF, PHYME. Ces résultats ainsi que l'annotation fonctionnelle des gènes étiquetés comme étant impliqués dans la réponse au stress seront intégrés dans l'interface de visualisation. Ce travail doit permettre une réflexion sur la notion d'orthologie fonctionnelle et effectuer une recherche translationnelle depuis les espèces modèles jusqu'aux espèces d'intérêt agronomique (i.e identifier des gènes candidats pour la réponse au stress du caféier à partir d'informations fonctionnelles connues chez Arabidopsis).

Summary

GenFam is an integrated gene families analysis system. This system allows (i) to create families of genes from complete genomes, (ii) to run a phylogenetic analysis of this family through the Galaxy workflow manager in order to define homology relationships (iii) to study evolutionary events from precomputed synteny blocks with the SynMap workflow of the Comparative Genomics platform (CoGe) and (iv) to integrate these results into a synthetic visualization interface. The first application of GenFam is to identify candidate genes in crops implicated in tolerance and adaptation mechanisms to environmental stress. This requires to be able to identify cis-regulatory elements which are specific to stress response (i.e. ABRE, DRE). In this context, we need to integrate new tools into GenFam in order to discover and search Transcription Factor Binding Sites (TFBS) in the promoter sequences of the family member genes. The Galaxy workflow, firstly, selects the flanking regions 5' or 3' to the genes of interest as selected by the user. Then, the flanking regions are analyzed with complementary methods such as MEME, STIF, PHYME. The results as well as gene functional annotations tagged as specific to stress response will be integrated into the GenFam visualization interface. This work should provide an insight into the concept of functional orthology and conduct translational research from model species to species of agronomic interest (i.e. identify candidate genes for stress response of the coffee from known functional information Arabidopsis).